

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

A bias corrected nonparametric regression estimator

Weixin Yao

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Yao, W. (2012). A bias corrected nonparametric regression estimator. Retrieved from <http://krex.ksu.edu>

Published Version Information

Citation: Yao, W. (2012). A bias corrected nonparametric regression estimator. *Statistics & Probability Letters*, 82(2), 274-282.

Copyright: © 2011 Elsevier B.V.

Digital Object Identifier (DOI): doi:10.1016/j.spl.2011.10.006

Publisher's Link: <http://www.sciencedirect.com/science/article/pii/S0167715211003270>

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

A Bias Corrected Nonparametric Regression Estimator

WEIXIN YAO

*Department of Statistics, The Kansas State University, Manhattan, Kansas 66506,
U.S.A.*

wxyao@ksu.edu

Abstract

In this article, we propose a new method of bias reduction in nonparametric regression estimation. The proposed new estimator has asymptotic bias order h^4 , where h is a smoothing parameter, in contrast to the usual bias order h^2 for the local linear regression. In addition, the proposed estimator has the same order of the asymptotic variance as the local linear regression. Our proposed method is closely related to the bias reduction method for kernel density estimate proposed by Chung and Lindsay (2011). However, our method is not a direct extension of their density estimate, but a totally new one based on the bias cancelation result of their proof.

Key Words: Bias reduction, Local linear regression, Nonparametric regression, Nonlinear smoother.

1 Introduction

Suppose the bivariate data $\{(x_1, y_1), \dots, (x_n, y_n)\}$ are from

$$Y = m(X) + \varepsilon,$$

where $E(\varepsilon \mid X) = 0$ and $\text{Var}(\varepsilon \mid X) = \sigma^2(X)$. Our main objective is to estimate the regression function $m(x)$. Nonparametric regression models provide flexible and powerful tools to estimate the regression function and are used in various fields. The generally used nonparametric regressions are: kernel, spline, local regression, and orthogonal series methods. See, for example, Fan and Gijbels (1996), Härdle (1990), and Wand and Jones

(1995). All the above nonparametric regression methods belong to linear smoothers, i.e., they have the form

$$\hat{m}(x_0) = \sum_{j=1}^n W_j(x_0, x_1, \dots, x_n) y_j, \quad (1.1)$$

where $W_j(\cdot)$ is the weight function and does not depend on (y_1, \dots, y_n) .

Among these methods, local linear regression (LLR) is one of the most popular and intuitive smoothing methods, due to both of its simplicity of computation and nice asymptotic properties. The LLR method finds $\boldsymbol{\theta} = (\beta_0, \beta_1)$ to minimize

$$\sum_{i=1}^n [K_h(x_i - x_0) \{y_i - \beta_0 - \beta_1(x_i - x_0)\}^2],$$

where $K_h(t) = h^{-1}K(t/h)$, h is the bandwidth, and $K(t)$ is a kernel function. Then the LLR estimate of $\hat{m}(x_0)$ is $\hat{\beta}_0$. It was shown that LLR has the asymptotic bias order of h^2 and is nearly 100% asymptotic minimax efficient among the class of linear smoothers. See Fan (1992, 1993) and Ruppert and Wand (1994), among others.

In this paper, we propose a new bias reduction nonparametric regression estimator, by employing the similar bias-correction idea of Chung and Lindsay (2011), which proposed a bias-corrected density estimator via a nonparametric mixture model using two-step EM algorithm. However, our new estimator is not a direct extension of their density estimator but a totally new one based on the bias cancelation idea of their proof. The new nonparametric regression estimator has the asymptotic bias order of h^4 and has the same order of asymptotic variance as LLR. Hence the new nonparametric regression is asymptotically more efficient than LLR and the asymptotic relative efficiency between the new method and LLR will go to infinity when $h \rightarrow 0$ or $n \rightarrow \infty$. Our new nonparametric regression method does not belong to the class of linear smoothers, which makes it possible for our method asymptotically more efficient than LLR. (Note that LLR is a nearly 100% asymptotic minimax efficient estimator among the class of linear smoothers.)

The idea of bias reduction has been widely used for kernel density estimation. Breiman, Meisel, and Purcell (1977) and Abramson (1982) proposed to use variable bandwidth to reduce the bias. Bartlett (1963) proposed to use the fourth-order kernels. Jones, Linton, and Nielsen (1995) considered a multiplicative bias correction method. DiMarzio and

Taylor (2004) applied boosting to the kernel density estimation. As Park et al. (1997) stated usually the extension of above bias reduction method to kernel based nonparametric regression is not entirely straightforward. However, there are some work, although much less, has been done on the bias reduction for nonparametric regression estimation. See, for example, Hall (1990), Linton and Nielsen (1994), Park et al. (1997), and Choi, Hall, and Rousson (2000).

The rest of the paper is organized as follows. Section 2 introduces our new nonparametric regression estimator and presents some asymptotic results. Some discussions and remarks are given in Section 4.

2 New nonparametric regression estimator

Let $f(x)$ be the density of the predictor X . Note that the usual kernel density estimator has the asymptotic bias order of h^2 . Chung and Lindsay (2011) proposed a new bias-corrected density estimate which has the asymptotic bias order of h^4 . The bias-corrected density estimator given (x_1, \dots, x_n) is

$$\hat{f}(x_0) = n^{-2}c(h) \sum_i \sum_j w_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0),$$

where $c(h) = 3\sqrt{2\pi}h$ and

$$w_j = \left\{ \frac{1}{n} \sum_{i=1}^n K_{\sqrt{2}h}(x_i - x_j) \right\}^{-1} \quad (2.1)$$

Here w_j is the regular kernel density estimator of $f(x_j)$ with bandwidth $\sqrt{2}h$.

Based on the bias-correction idea of their proof, we propose the following bias-corrected nonparametric regression estimate (BRNP)

$$\hat{m}(x_0) = \frac{n^{-2}c(h) \sum_i \sum_j v_j y_i y_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0)}{n^{-2}c(h) \sum_i \sum_j w_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0)} \equiv \frac{N}{D} \quad (2.2)$$

where

$$v_j = \left\{ \frac{1}{n} \sum_{i=1}^n K_{\sqrt{2}h}(x_i - x_j) y_i \right\}^{-1}, \quad (2.3)$$

$c(h) = 3\sqrt{2\pi}h$, and w_j is as defined in (2.1). For simplicity of calculation, we will use Gaussian kernel for $K(\cdot)$ in this paper. Note that the new regression estimator (2.2) does not belong to the class of linear smoothers (defined in (1.1)). The denominator D in (2.2) is the bias reduction density estimator of $f(x)$, proposed by Chung and Lindsay (2011), which has the asymptotic bias order of h^4 . Similar to the proof of D in Chung and Lindsay (2011), we can check that the nominator N also has the asymptotic bias order of h^4 for estimating $m(x)f(x)$. Hence the proposed estimator (2.2) has the asymptotic bias order of h^4 for estimating $m(x)$.

Theorem 2.1. *Let x_0 be an interior point of the support of X . Under regularity conditions A1–A4 in the appendix, for the estimate $\hat{m}(x_0)$ defined in (2.2), we have*

$$\text{Bias}\{\hat{m}(x_0)\} = \frac{h^4\{A - m(x_0)B\}}{f(x_0)} + o(h^4), \quad (2.4)$$

and

$$\text{Var}\{\hat{m}(x_0)\} = \frac{\sigma^2(x_0)}{nh\sqrt{\pi}f(x_0)} \left(\sqrt{2} + \frac{1}{4} - \frac{2}{\sqrt{3}} \right) + o\left(\frac{1}{nh}\right) \approx \frac{0.2875\sigma^2(x_0)}{nhf(x_0)} + o\left(\frac{1}{nh}\right), \quad (2.5)$$

where

$$A = [-g^{(4)}(x_0) + g^{-1}(x_0)g'(x_0)g'''(x_0) + g^{-1}(x_0)g''(x_0)^2 - g'(x_0)^2g^{-2}(x_0)g''(x_0)] \quad (2.6)$$

and

$$B = [-f^{(4)}(x_0) + f^{-1}(x_0)f'(x_0)f'''(x_0) + f^{-1}(x_0)f''(x_0)^2 - f'(x_0)^2f^{-2}(x_0)f''(x_0)], \quad (2.7)$$

with $g(x) = m(x)f(x)$.

The proof of Theorem 2.1 is given in the appendix. From the above theorem, we can see that the asymptotic bias order of new method is h^4 compared to the asymptotic bias order of h^2 for LLR. Note, however, the bias term in (2.4) depends on the design density

$f(x_0)$. Therefore, the proposed bias reduction nonparametric regression is not adaptive to nonuniform designs. In addition, it can be seen that the asymptotic variance of LLR is

$$\frac{0.2821\sigma^2(x_0)}{nhf(x_0)} + o\left(\frac{1}{nh}\right),$$

when the Gaussian kernel is used. Hence the asymptotic variance of the new method has the same order as LLR and is almost the same as LLR when the bandwidth is the same. So when $n \rightarrow \infty$ and $h \rightarrow 0$, the asymptotic relative efficiency between our new method and LLR will go to infinity.

Bandwidth selection. Note that the mean squared error (MSE) of the $\hat{m}(x_0)$ in (2.2) is

$$\text{Bias}\{\hat{m}(x_0)\} + \text{Var}\{\hat{m}(x_0)\}.$$

Based on Theorem 2.1, the asymptotic optimal bandwidth that minimizes the asymptotic MSE, is

$$h_{opt}(x) = \left[\frac{0.036\sigma^2(x_0)f(x_0)}{n\{A - m(x_0)B\}^2} \right]^{1/9} \quad (2.8)$$

where A and B are as defined in (2.6) and (2.7).

The asymptotic global optimal bandwidth, which minimizes the asymptotic weighted Mean Integrated Square Error $\int [\widehat{\text{Bias}}\{\hat{m}(x)\} + \widehat{\text{Var}}\{\hat{m}(x)\}] w(x)dx$, where $w \geq 0$ is some weight function, is

$$h_{opt} = \left[\frac{0.036 \int \sigma^2(x)w(x)/f(x)dx}{n \int \{A - m(x)B\}^2 w(x)/f(x)^2 dx} \right]^{1/9}, \quad (2.9)$$

where A and B are as defined in (2.6) and (2.7).

To select the bandwidth for our new method in practice, one can replace the unknown quantities in (2.8) and (2.9) by their estimates. Another method is to use leave-one out cross-validation to select the bandwidth h , i.e., we find h by minimizing

$$CV(h) = \sum_{i=1}^n \{y_i - \hat{m}_{(i,h)}(x_i)\}^2 \quad (2.10)$$

where $\hat{m}_{(i,h)}(\cdot)$ is the proposed regression estimate by leaving i th observation out.

Computational issue. Notice that, by symmetry, we have

$$\begin{aligned} N &= n^{-2}c(h) \sum_i \sum_j \frac{v_i + v_j}{2} y_i y_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0) \\ &= \left(2 \sum_{j=1}^n \sum_{i=1}^{j-1} + \sum_{i=j}^n \right) \frac{v_i + v_j}{2} y_i y_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0), \end{aligned}$$

where N is the nominator of (2.2). Similarly, by symmetry, we also have

$$\begin{aligned} D &= n^{-2}c(h) \sum_i \sum_j \frac{w_i + w_j}{2} K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0) \\ &= \left(2 \sum_{j=1}^n \sum_{i=1}^{j-1} + \sum_{i=j}^n \right) \frac{w_i + w_j}{2} K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0), \end{aligned}$$

where D is the denominator of (2.2). One can then replace the full double summation in (2.2) by using $2 \sum_{j=1}^n \sum_{i=1}^{j-1} + \sum_{i=j}^n$, which cuts the number of summands to compute nearly in half.

Note that v_j defined in (2.3) converges to $\{m(x_j)f(x_j)\}^{-1}$. Hence if $m(x_j)$ is close to 0, there will be some numerical issues to calculate v_j^{-1} . In order to reduce the computational uncertainty, we recommend to add (or minus) a large enough positive constant to all the response to make all the transformed response well above (or below) zero when the original response y 's change signs. For example, suppose we add a constant A to all the response y 's, then the regression estimate, say $\hat{m}^*(x)$, based on the transformed response, minus A will be our proposed regression estimate $\hat{m}(x)$ for the original response. In practice, we can simply let $A = c + \max\{\text{abs}(y_1), \dots, \text{abs}(y_n)\}$, for some constant c , if y 's change signs.

Boundary properties. It is well known that the traditional kernel density estimate has the boundary effect. Similarly, the the bias reduction density estimate of Chung and Lindsay (2011) and the proposed nonparametric regression estimator also have boundary effect. Suppose the support of X has one or more known boundaries. Without loss of generality, let us assume that the support of X is $[0, \infty)$ and consider the boundary points $x_b = b\sqrt{2}h, 0 \leq b < 1$ (note that the bandwidth $\sqrt{2}h$ is used for w_j in (2.1)). Then, the bias reduction density estimate $\hat{f}(x_b)$ of Chung and Lindsay (2011) has expectation $f(x_b) \int_{-\infty}^b K(t)dt + O(h)$ and N in (2.2) has expectation $m(x_b)f(x_b) \int_{-\infty}^b K(t)dt + O(h)$.

Therefore, the proposed nonparametric regression estimator $\hat{m}(x_b)$ has expectation $m(x_b) + O(h)$. Note that for the boundary point x_b , the bias reduction density estimate $\hat{f}(x_b)$ of Chung and Lindsay (2011) is not consistent but $\hat{m}(x_b)$ is still consistent but with bias order of $O(h)$. Many methods have been proposed to do boundary correction. See, for example, Müller (1991), Jones (1993), Marron and Ruppert (1994), and Zhang, Karunamuni, and Jones (1999). It will be interesting to see whether the above methods can be extended to the kernel density estimate of Chung and Lindsay (2011) and our proposed bias reduction nonparametric regression estimate. These require further research.

3 Simulation study

In this section, we will conduct a Monte Carlo simulation study to assess the finite sample performance of the proposed method and compare it with LLR. Note that v_j defined in (2.3) converges to $\{m(x_j)f(x_j)\}^{-1}$. Hence if $m(x_j)$ is close to 0, it is numerically not stable to calculate v_j^{-1} . In order to reduce the computational uncertainty, we recommend to add (or minus) a large enough positive constant to all the response to make all the transformed response well above (or below) zero when the original response y 's change signs. For example, suppose we add a constant A to all the response y 's, then the regression estimate, say $\hat{m}^*(x)$, based on the transformed response, minus A will be our proposed regression estimate $\hat{m}(x)$ for the original response. In practice, we can simply let $A = c + \max\{\text{abs}(y_1), \dots, \text{abs}(y_n)\}$, for some constant c , if y 's change signs.

Example: We generate the independent and identically distributed (i.i.d.) data $\{(x_i, y_i), i = 1, \dots, n\}$ from the model

$$Y_i = 10 + 2 \sin(\pi X_i) + \epsilon_i,$$

where $X_i \sim N(0, 1)$ and $\epsilon_i \sim N(0, 1)$.

We compared the following five estimators:

1. Local linear regression (LLR) with the plug-in bandwidth (Ruppert, Sheather, and Wand, 1995).

2. Local linear regression with the asymptotic global optimal bandwidth assuming the whole generation model is known (LLR_{opt}).
3. Local linear regression with the bandwidth chosen by cross validation (LLR_{CV}).
4. Our proposed bias reduction estimator with the asymptotic global optimal bandwidth assuming the whole generation model is known ($BRNP_{opt}$).
5. The proposed bias reduction estimator with bandwidth chosen by cross validation ($BRNP_{CV}$).

In order to compare different methods, we reported the relative efficiency between (LLR_{opt} , LLR_{CV} , $BRNP_{opt}$, $BRNP_{CV}$) and LLR in Table 1. For example, $RE(LLR_{opt})$ is the relative efficiency between the LLR_{opt} estimator and the LLR estimator based on 100 grids points (t_1, \dots, t_{100}) equally spaced from $(-2, 2)$ based on 500 replications, i.e.,

$$RE(LLR_{opt}) = \frac{MSE(LLR)}{MSE(LLR_{opt})}$$

where

$$MSE(LLR) = \frac{1}{500} \frac{1}{100} \sum_{s=1}^{500} \sum_{j=1}^{100} \{\hat{m}_s^{LLR}(t_j) - m(t_j)\}^2$$

and

$$MSE(LLR_{opt}) = \frac{1}{500} \frac{1}{100} \sum_{s=1}^{500} \sum_{j=1}^{100} \{\hat{m}_s^{LLR_{opt}}(t_j) - m(t_j)\}^2,$$

where $\hat{m}_s^{LLR}(t_j)$ is the LLR estimate of $m(t_j)$ based on s th replication and $\hat{m}_s^{LLR_{opt}}(t_j)$ is the LLR_{opt} estimate of $m(t_j)$ based on s th replication. The same notation applies to other methods. In the right panel of Table 1, we also reported the relative efficiency between $BRNP_{opt}$ and LLR_{opt} (denoted by $RE(OPT)$) and the relative efficiency between $BRNP_{CV}$ and LLR_{CV} (denoted by $RE(CV)$): a value greater than 1 indicates that the proposed method performs better than the LLR.

From the Table 1, we can see that when sample size increases from 200 to 1600 the relative efficiency $RE(OPT)$ and $RE(CV)$ are also increasing, which is expected based on the asymptotic results in Section 2. Surprisingly, our proposed estimate BRNP works even better than LLR for small sample size 50 and 100.

Table 1: Relative efficiency between different estimators and the LLR estimator

n	RE(LLR _{opt})	RE(LLR _{CV})	RE(BRNP _{opt})	RE(BRNP _{CV})	RE(OPT)	RE(CV)
50	1.7374	1.8997	3.3161	3.2857	1.9087	1.7305
100	1.7903	1.8785	2.2553	2.2593	1.2597	1.2027
200	1.5347	1.5411	1.7097	1.7112	1.1140	1.1104
400	1.2711	1.2719	1.4529	1.4470	1.1430	1.1377
800	1.3305	1.3040	1.6110	1.5929	1.2108	1.2215
1600	1.1271	1.1100	1.4288	1.4216	1.2872	1.2834

We should point out, though, the relative performance between LLR and BRNP depends on the true regression function and the design density of the predictor (note that the asymptotic bias of LLR does not depend on the design density but ours does). Although the BRNP has higher order of asymptotic bias term, the multiplicative constant $\{A - m(x_0)B\}/f(x_0)$ in (2.4) is very complicated and could make the bias term even larger (and thus larger MSE, due to similar asymptotic variance) than the simpler bias term of LLR, for finite sample size when h is not small enough.

4 Discussion

It is of interest to extend our new method to other regression models such as partially linear model. This requires more research in future. One might also directly use the bias corrected density estimator of Chung and Lindsay (2011) for the multivariate variable (X, Y) . Then, we can estimate the conditional expectation $E(Y \mid X = x)$ based on the estimated conditional density of Y given X . Denote by $\tilde{m}(x)$ the derived estimate of $m(x)$. By some calculation, we can see that $\tilde{m}(x)$ involves both the bandwidth for X and the bandwidth for Y and the estimation formula is much complicated than our proposed new estimator (2.2). Therefore, the choice of bandwidths will be much harder and it is desirable to remove the bandwidth dependence for Y . Some further research is needed in this direction.

5 Appendix: Proofs

The following technical conditions are imposed in this section. They are not the weakest possible conditions, but they are imposed to facilitate the proofs

(A1) $f^{(5)}(x)$ is a bounded function and continuous at the point x_0 , where x_0 is an interior point of the support of X and $f^{(5)}(\cdot)$ is the fifth derivative of $f(\cdot)$.

(A2) $m^{(5)}(x)$ is a bounded function and continuous at the point x_0 , where $m(x) = E(Y | X = x)$ is the regression function.

(A3) $\sigma^{(5)}(x)$ is a bounded function and continuous at the point x_0 , where $\sigma^2(x) = \text{Var}(Y | X = x)$ is the variance function.

(A4) $m(x_0) \neq 0$ and $f(x_0) \neq 0$.

Proof of Theorem 2.1:

Let

$$N \equiv n^{-2}c(h) \sum_i \sum_j v_j y_i y_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0) \quad (5.1)$$

$$D \equiv n^{-2}c(h) \sum_i \sum_j w_j K_{\sqrt{3}h}(x_i - x_j) K_{\sqrt{3}h}(x_i - x_0) K_{\sqrt{3}h}(x_j - x_0). \quad (5.2)$$

We have

$$\begin{aligned} N &\rightarrow \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2}) k_{2t}(x_j, x_0) y_i y_j f(x_i, y_i) dx_i dy_i}{\int k_{2t}(x_k, x_j) y_k f(x_k, y_k) dx_k dy_k} f(x_j, y_j) dx_j dy_j \\ &= \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2}) k_{2t}(x_j, x_0) m(x_i) f(x_i) dx_i}{\int k_{2t}(x_k, x_j) m(x_k) f(x_k) dx_k} m(x_j) f(x_j) dx_j, \end{aligned}$$

where $t = h^2$, $k_t(x_i, x_j) = K_h(x_i - x_j)$, and $K_h(\cdot)$ is the normal density with mean 0 and variance h^2 .

Let

$$x_i = \frac{x_0 + x_j}{2} + \sqrt{3/2}h z_1, \quad x_k = x_j + \sqrt{2}h z_2, \quad x_j = x_0 + \sqrt{2}h z_3.$$

Then

$$N \rightarrow \int \frac{\int \phi(z_1)g(x_0 + \frac{\sqrt{2}}{2}hz_3 + \sqrt{\frac{3}{2}}hz_1)dz_1}{\int \phi(z_2)g(x_0 + \sqrt{2}hz_3 + \sqrt{2}hz_2)dz_2} \phi(z_3)g(x_0 + \sqrt{2}hz_3)dz_3$$

where, $g(x) = m(x)f(x)$. Then

$$\begin{aligned} & \int \phi(z_1)g(x_0 + \frac{\sqrt{2}}{2}hz_3 + \sqrt{\frac{3}{2}}hz_1)dz_1 \\ &= g(x_0) + \frac{\sqrt{2}}{2}g'(x_0)hz_3 + \frac{g''(x_0)}{4}h^2(z_3^2 + 3) \\ & \quad + \frac{\sqrt{2}g'''(x_0)}{4!}h^3(z_3^3 + 9z_3) + \frac{g^{(4)}(x_0)}{4!}h^4 \left(\frac{1}{4}z_3^4 + \frac{27}{4} + \frac{9}{2}z_3^2 \right) + o(h^4), \end{aligned}$$

and

$$\begin{aligned} & \frac{g(x_0 + \sqrt{2}hz_3)}{\int \phi(z_2)g(x_0 + \sqrt{2}hz_2 + \sqrt{2}hz_3)dz_2} \\ &= 1 - g^{-1}g''h^2 - \sqrt{2}h^3z_3g^{-1}(g''' - g^{-1}g'g'') + h^4g^{-1}\{2g^{-1}g'z_3^2g''' \\ & \quad + g^{-1}(g'')^2z_3^2 + g^{-1}(g'')^2 - 2g^{-2}(g')^2z_3^2g'' - \frac{g^{(4)}}{2} - g^{(4)}z_3^2\} + o(h^4). \end{aligned}$$

Hence

$$\begin{aligned} N &\rightarrow g(x_0) + h^4 [-g^{(4)}(x_0) + g^{-1}(x_0)g'(x_0)g'''(x_0) + g^{-1}(x_0)g''(x_0)^2 - g'(x_0)^2g^{-2}(x_0)g''(x_0)] + o(h^4) \\ &= g(x_0) + h^4A + o(h^4) \end{aligned}$$

similarly, we can get

$$\begin{aligned} D &\rightarrow f(x_0) + h^4 [-f^{(4)}(x_0) + f'(x_0)f'''(x_0)/f(x_0) + f''(x_0)^2/f(x_0) - f'(x_0)^2f''(x_0)/f^2(x_0)] + o(h^4) \\ &= f(x_0) + h^4B + o(h^4), \end{aligned}$$

where A and B are given in (2.6) and (2.7).

Then

$$\hat{m}(x_0) \rightarrow \frac{g(x_0) + h^4A + o(h^4)}{f(x_0) + h^4B + o(h^4)} = m(x_0) + \frac{h^4\{A - m(x_0)B\}}{f(x_0)} + o(h^4).$$

The asymptotic bias result (2.4) follows directly.

In order to find the asymptotic variance of the proposed estimator (2.2), we rewrite the estimator as a functional on the distribution $T(\hat{F})$,

$$T(\hat{F}) = \frac{\int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y_i y_j d\hat{F}(x_i, y_i)}{\int k_{2t}(x_k, x_j)y_k d\hat{F}(x_k, y_k)} d\hat{F}(x_j, y_j)}{\int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)d\hat{F}(x_i, y_i)}{\int k_{2t}(x_k, x_j)d\hat{F}(x_k, y_k)} d\hat{F}(x_j, y_j)} \equiv \frac{A(\hat{F})}{B(\hat{F})} \quad (5.3)$$

To do the calculation, we will find the first von Mises derivative $T'(x, y)$, and then use the result that

$$T(\hat{F}) - T(F) \approx \int T'(x, y)d(\hat{F} - F) \quad (5.4)$$

so that the asymptotic variance of $T(\hat{F})$ is

$$asyvar(T(\hat{F})) = \text{Var}_F(T'(x, y))/n.$$

The first von Mises derivative $A'(x, y)$ is:

$$\begin{aligned} A'(x, y) &= \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y_i y_j dF(x_i, y_i)}{\int k_{2t}(x_k, x_j)y_k dF(x_k, y_k)} d\Delta(x_j, y_j) \\ &+ \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y_i y_j d\Delta(x_i, y_i)}{\int k_{2t}(x_k, x_j)y_k dF(x_k, y_k)} dF(x_j, y_j) \\ &- \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y_i y_j dF(x_i, y_i) \int k_{2t}(x_k, x_j)y_k d\Delta(x_k, y_k)}{(\int k_{2t}(x_k, x_j)y_k dF(x_k, y_k))^2} dF(x_j, y_j) \end{aligned}$$

The first von Mises derivative $B'(x, y)$ is:

$$\begin{aligned} B'(x, y) &= \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)dF(x_i, y_i)}{\int k_{2t}(x_k, x_j)dF(x_k, y_k)} d\Delta(x_j, y_j) \\ &+ \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)d\Delta(x_i, y_i)}{\int k_{2t}(x_k, x_j)dF(x_k, y_k)} dF(x_j, y_j) \\ &- \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)dF(x_i, y_i) \int k_{2t}(x_k, x_j)d\Delta(x_k, y_k)}{(\int k_{2t}(x_k, x_j)dF(x_k, y_k))^2} dF(x_j, y_j) \end{aligned}$$

where the measure $d\Delta(x_j, y_j) = d\delta_{(x,y)} - dF(x_j, y_j)$ and $\delta_{(x,y)}$ is the distribution degenerate at (x, y) . Then

$$T'(x, y) = \frac{A'(x, y)B(F) - A(F)B'(x, y)}{B^2(F)} = \frac{A'(x, y)}{B(F)} - \frac{A(F)}{B(F)^2}B'(x, y). \quad (5.5)$$

We can rewrite by letting

$$\begin{aligned} I_1(x, y) &= B(F)^{-1} \left(\frac{\int k_{3t_1/2}(x_i, \frac{x_0+x}{2})k_{2t}(x, x_0)y_i y dF(x_i, y_i)}{\int k_{2t}(x_k, x)y_k dF(x_k, y_k)} \right) \\ &\quad - \frac{A(F)}{B(F)^2} \left(\frac{\int k_{3t/2}(x_i, \frac{x_0+x}{2})k_{2t}(x, x_0)dF(x_i, y_i)}{\int k_{2t}(x_k, x)dF(x_k, y_k)} \right) \\ I_2(x, y) &= B(F)^{-1} \left(\int \frac{k_{3t/2}(x, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y y_j dF(x_j, y_j)}{\int k_{2t}(x_k, x_j)y_k dF(x_k, y_k)} \right) \\ &\quad - \frac{A(F)}{B(F)^2} \left\{ \int \frac{k_{3t/2}(x, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)}{\int k_{2t}(x_k, x_j)dF(x_k, y_k)} dF(x_j, y_j) \right\} \\ I_3(x, y) &= -B(F)^{-1} \left(\int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t}(x_j, x_0)y_i y_j dF(x_i, y_i)k_{2t_1}(x, x_j)y}{\left(\int k_{2t}(x_k, x_j)y_k dF(x_k, y_k)\right)^2} dF(x_j, y_j) \right) \\ &\quad + \frac{A(F)}{B(F)^2} \left\{ \int \frac{\int k_{3t/2}(x_i, \frac{x_0+x_j}{2})k_{2t_1}(x_j, x_0)dF(x_i, y_i)k_{2t}(x, x_j)}{\left(\int k_{2t}(x_k, x_j)dF(x_k, y_k)\right)^2} dF(x_j, y_j) \right\} \end{aligned}$$

and noting that

$$T'(x, y) = I_1(x, y) + I_2(x, y) + I_3(x, y) - E(I_1 + I_2 + I_3)$$

Thus we seek $\text{Var}(I_1 + I_2 + I_3)$. Since $E(I_1 + I_2 + I_3) = 0$, we only need to find $E(I_1 + I_2 + I_3)^2$.

We will prove the following limiting results

$$h \int I_1^2(x, y) f(x, y) dx dy \rightarrow \frac{\sigma^2(x_0)}{2\sqrt{2\pi}f(x_0)} \quad (5.6)$$

$$h \int I_2^2(x, y) f(x, y) dx dy \rightarrow \frac{\sigma^2(x_0)}{2\sqrt{2\pi}f(x_0)} \quad (5.7)$$

$$h \int I_3^2(x, y) f(x, y) dx dy \rightarrow \frac{\sigma^2(x_0)}{4\sqrt{\pi}f(x_0)} \quad (5.8)$$

$$h \int I_1(x, y) I_2(x, y) f(x, y) dx dy \rightarrow \frac{\sigma^2(x_0)}{2\sqrt{2\pi}f(x_0)} \quad (5.9)$$

$$h \int I_1(x, y) I_3(x, y) f(x, y) dx dy \rightarrow -\frac{\sigma^2(x_0)}{2\sqrt{3\pi}f(x_0)} \quad (5.10)$$

$$h \int I_2(x, y) I_3(x, y) f(x, y) dx dy \rightarrow -\frac{\sigma^2(x_0)}{2\sqrt{3\pi}f(x_0)} \quad (5.11)$$

For simplicity, we will only prove the first result. The other results follow the similar arguments.

Note that

$$\begin{aligned} B(F) &\rightarrow \int f(x_0, y) dy = f(x_0); \\ A(F) &\rightarrow \int y f(x_0, y) dy = m(x_0) f(x_0); \end{aligned}$$

where $m(x_0) = \int y f(y | x_0) dy$.

Let

$$\begin{aligned} x_i &= \frac{x_0 + x}{2} + \sqrt{3/2} h z_1 \\ x_k &= x + \sqrt{2} h z_2 \\ x &= x_0 + \sqrt{2} h z_3. \end{aligned}$$

Then,

$$\begin{aligned}
& \int \left(\frac{\int k_{3t/2}(x_i, \frac{x_0+x}{2}) k_{2t}(x, x_0) y_i y dF(x_i, y_i)}{\int k_{2t}(x_k, x) y_k dF(x_k, y_k)} \right)^2 f(x, y) dx dy \\
&= \frac{1}{\sqrt{2}h} \int \left(\frac{\int \phi(z_1) \phi(z_3) y_i y f(x_0 + \frac{\sqrt{2}}{2} h z_3 + \sqrt{\frac{3}{2}} h z_1, y_i) dz_1 dy_i}{\int \phi(z_2) y_k f(x_0 + \sqrt{2} h z_3 + \sqrt{2} h z_2, y_k) dz_2 dy_k} \right)^2 f(x_0 + \sqrt{2} h z_3, y) dz_3 dy \\
&= \frac{\{m^2(x_0) + \sigma^2(x_0)\} f(x_0)}{\sqrt{2}h} \int \phi^2(z_3) dz_3 (1 + o(1)) \\
&= \frac{\{m^2(x_0) + \sigma^2(x_0)\} f(x_0)}{2\sqrt{2\pi}h} (1 + o(1)).
\end{aligned}$$

by noting that

$$\int \phi^2(z) dz = k_2(0, 0) = \frac{1}{2\sqrt{\pi}}.$$

Using similarly computation, we can get

$$\begin{aligned}
& \int \left(\frac{\int k_{3t/2}(x_i, \frac{x_0+x}{2}) k_{2t}(x, x_0) y_i y dF(x_i, y_i)}{\int k_{2t}(x_k, x) y_k dF(x_k, y_k)} \right) \left(\frac{\int k_{3t/2}(x_i, \frac{x_0+x}{2}) k_{2t}(x, x_0) dF(x_i, y_i)}{\int k_{2t}(x_k, x) dF(x_k, y_k)} \right) f(x, y) dx dy \\
&= \int \left(\frac{\int \phi(z_1) \phi(z_3) y_i y f(x_0 + \frac{\sqrt{2}}{2} h z_3 + \sqrt{\frac{3}{2}} h z_1, y_i) dz_1 dy_i}{\int \phi(z_2) y_k f(x_0 + \sqrt{2} h z_3 + \sqrt{2} h z_2, y_k) dz_2 dy_k} \right) \\
&\quad \times \left(\frac{\int \phi(z_1) \phi(z_3) f(x_0 + \frac{\sqrt{2}}{2} h z_3 + \sqrt{\frac{3}{2}} h z_1, y_i) dz_1 dy_i}{\int \phi(z_2) f(x_0 + \sqrt{2} h z_3 + \sqrt{2} h z_2, y_k) dz_2 dy_k} \right) f(x_0 + \sqrt{2} h z_3, y) dz_3 dy \\
&= \frac{m(x_0) f(x_0)}{2\sqrt{2\pi}h} (1 + o(1))
\end{aligned}$$

and

$$\begin{aligned}
& \int \left(\frac{\int k_{3t/2}(x_i, \frac{x_0+x}{2}) k_{2t}(x, x_0) dF(x_i, y_i)}{\int k_{2t}(x_k, x) dF(x_k, y_k)} \right)^2 f(x, y) dx dy \\
&= \int \left(\frac{\int \phi(z_1) \phi(z_3) f(x_0 + \frac{\sqrt{2}}{2} h z_3 + \sqrt{\frac{3}{2}} h z_1, y_i) dz_1 dy_i}{\int \phi(z_2) f(x_0 + \sqrt{2} h z_3 + \sqrt{2} h z_2, y_k) dz_2 dy_k} \right)^2 f(x_0 + \sqrt{2} h z_3, y) dz_3 dy \\
&= \frac{1}{2\sqrt{2\pi}h} f(x_0) (1 + o(1)).
\end{aligned}$$

Hence,

$$h \int I_1^2(x, y) f(x, y) dx dy \rightarrow \frac{\sigma^2(x_0)}{2\sqrt{2\pi}f(x_0)}.$$

By similar arguments we can get all other results and hence the formula (2.5).

6 Acknowledgements

The authors are grateful to the editor and the referee for their insightful comments and suggestions, which greatly improved this article. The author would also like to thank Bruce G. Lindsay and Yeojin Chung for their valuable discussions and advice in this research, as well as their generosity in providing the manuscript of Chung and Lindsay (2011).

References

- Abramson, I. S. (1982). On bandwidth variation in kernel estimates—a square root law. *Annals of Statistics*, 10, 1217-1223.
- Barlett, M. S. (1963). Statistical estimation of density function. *Sankhya*, Ser. A, 15, 245-254.
- Breiman, L., Meisel, W., and Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, 19, 135-144.
- Choi, E., Hall, P., and Rousson, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Annals of Statistics*, 28, 1339-1355.
- Chung, Y. and Lindsay, B. G. (2011). A Likelihood-tuned density estimator via a nonparametric mixture model. In *Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger*, ed. D.R. Hunter, D.St.P. Richards, and J.L. Rosenberger, 69-89. World Scientific Publishing.
- DiMarzio, M. and Taylor, C. (2004). Boosting kernel density estimates: A bias reduction technique? *Biometrika*, 91, 226-233.

- Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87, 998-1004.
- Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, 87, 998-1004.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Hall, P. (1990). On the bias of variable bandwidth curve estimators. *Biometrika*, 77, 529-535.
- Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, 3, 135-146.
- Jones, M. C., Linton, O., and Nielsen, J. P. (1995). A simple and effective bias reduction method for kernel density estimation. *Biometrika*, 82, 327-338.
- Linton, O. and Nielsen, J. P. (1994). A multiplicative bias reduction method for nonparametric regression. *Statistics and Probability Letter*, 19, 181-187.
- Marron, J. S., and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society, Ser. B*, 56, 653-671.
- Muller, H. G. (1991). Smooth optimum kernel estimators near end-points. *Biometrika*, 78, 521-530.
- Park, B. U., Kim, W. C., Ruppert, D., Jones, M. C., Signorini, D. F., and Kohn, R. (1997). Simple transformation techniques for improved nonparametric regression. 24, 145-163.
- Ruppert, D., Sheather S. J., and Wand M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of American Statistical Association*, 90, 1257-1270.
- Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics*, 22, 1346-1370.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*, London: Chapman and Hall.

Zhang, S., Karunamuni, R. J., and Jones, M. C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, 94, 1231-1241.